

# Constructing Adjacency Arrays from Incidence Arrays

Hayden Jananthan<sup>1,2</sup> Karia Dibert<sup>2,3</sup> Jeremy Kepner<sup>2,3,4</sup>

<sup>1</sup>Vanderbilt University Mathematics Department, <sup>2</sup>MIT Lincoln Laboratory Supercomputing Center,

<sup>3</sup>MIT Mathematics Department, <sup>4</sup>MIT Computer Science & AI Laboratory

**Abstract**—Graph construction, a fundamental operation in a data processing pipeline, is typically done by multiplying the incidence array representations of a graph,  $\mathbf{E}_{\text{in}}$  and  $\mathbf{E}_{\text{out}}$ , to produce an adjacency array of the graph,  $\mathbf{A}$ , that can be processed with a variety of algorithms. This paper provides the mathematical criteria to determine if the product  $\mathbf{A} = \mathbf{E}_{\text{out}}^T \mathbf{E}_{\text{in}}$  will have the required structure of the adjacency array of the graph. The values in the resulting adjacency array are determined by the corresponding addition  $\oplus$  and multiplication  $\otimes$  operations used to perform the array multiplication. Illustrations of the various results possible from different  $\oplus$  and  $\otimes$  operations are provided using a small collection of popular music metadata.

**Keywords**—graph; incidence array; adjacency array; semiring

## I. INTRODUCTION

The duality between the canonical representation of graphs as abstract collections of vertices and edges and a matrix representation has been a part of graph theory since its inception [Konig 1931], [Konig 1936]. Matrix algebra has been recognized as a useful tool in graph theory for nearly as long [Harary 1969], [Sabadusi 1960], [Weischel 1962], [McAndrew 1963], [Teh & Yap 1964], [McAndrew 1965], [Harary & Tauth 1964], [Brualdi 1967]. The modern description of the duality between graph algorithms and matrix mathematics (or sparse linear algebra) has been extensively covered in the recent literature [Kepner & Gilbert 2011] and has further spawned the development of the GraphBLAS math library standard (GraphBLAS.org)[Mattson et al 2013] that has been developed in a series of proceedings [Mattson 2014a], [Mattson 2014b], [Mattson 2015], [Buluç 2015], [Mattson 2016] and implementations [Buluç & Gilbert 2011], [Kepner et al 2012], [Ekanadham et al 2014], [Hutchison et al 2015], [Anderson et al 2016], [Zhang et al 2016].

Adjacency arrays, typically denoted  $\mathbf{A}$ , have much in common with adjacency matrices. Likewise, incidence arrays or edge arrays, typically denoted  $\mathbf{E}$ , have much in common with incidence matrices [Bruck & Ryser 1949], [Ford & Fulkerson 1962],

[Fulkerson & Gross 1965], [Fisher & Wing 1965], edge matrices [Dobrzanskyj & Freudenstein 1967], adjacency lists [Bodin & Kursh 1979], and adjacency structures [Tarjan 1972]. The powerful link between adjacency arrays and incidence arrays via array multiplication is the focus of the first part of this paper.

Incidence arrays are often readily obtained from raw data. In many cases, an associative array representing a spreadsheet or database table is already in the form of an incidence array. However, to analyze a graph, it is often convenient to represent the graph as an adjacency array. Constructing an adjacency array from data stored in an incidence array via array multiplication is one of the most common and important steps in a data processing system.

Given a graph  $G$  with vertex set  $K_{\text{out}} \cup K_{\text{in}}$  and edge set  $K$ , the construction of adjacency arrays for  $G$  relies on the assumption that  $\mathbf{E}_{\text{out}}^T \mathbf{E}_{\text{in}}$  is an adjacency array of  $G$ . This assumption is certainly true in the most common case where the value set is composed of non-negative reals and the operations  $\oplus$  and  $\otimes$  are arithmetic plus (+) and arithmetic times ( $\times$ ) respectively. However, one hallmark of associative arrays is their ability to contain as values nontraditional data. For these value sets,  $\oplus$  and  $\otimes$  may be redefined to operate on non-numerical values. For example, for the value of all alphanumeric strings, with

$$\oplus = \max()$$

$$\otimes = \min()$$

it is not immediately apparent in this case whether  $\mathbf{E}_{\text{out}}^T \mathbf{E}_{\text{in}}$  is an adjacency array of the graph whose set of vertices is  $K_{\text{out}} \cup K_{\text{in}}$ . In the subsequent sections, the criteria on the value set  $V$  and the operations  $\oplus$  and  $\otimes$  are presented so that

$$\mathbf{A} = \mathbf{E}_{\text{out}}^T \mathbf{E}_{\text{in}}$$

always produces an adjacency array [Dibert et al 2015].

## A. Definitions

For a directed graph (from here onwards, just ‘graph’)  $G$ ,  $K_{\text{out}}$  will denote the set of vertices which are the sources of edges,  $K_{\text{in}}$  will denote the set of vertices which are the targets of edges, and  $K$  will denote the set of edges. The vertex set of  $G$  will be assumed to be  $K_{\text{out}} \cup K_{\text{in}}$ .  $K_{\text{out}}$ ,  $K_{\text{in}}$ , and  $K$  are assumed to be finite and totally-ordered.

This material is based in part upon work supported by the NSF under grant number DMS-1312831. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

$V$  will denote the set of values that the data can take on, such as non-negative real numbers or the elements of an ordered set.  $\oplus$  and  $\otimes$  are binary operations on  $V$  (in particular,  $V$  is closed under the operations  $\oplus$  and  $\otimes$ ), such as  $\oplus = +$  and  $\otimes = \times$  or  $\oplus = \max$  and  $\otimes = +$ .  $\oplus$  and  $\otimes$  each have identity elements 0 and 1, respectively, i.e.

$$\begin{aligned} v \oplus 0 &= 0 \oplus v = v \\ v \otimes 1 &= 1 \otimes v = v \end{aligned}$$

for all  $v \in V$ .

For the purposes of understanding what algebraic properties are required for  $\mathbf{E}_{\text{out}}^\top \mathbf{E}_{\text{in}}$  to be an adjacency array of a graph,  $\oplus$  and  $\otimes$  will not be assumed to be associative or commutative, and  $\otimes$  does not necessarily distribute over  $\oplus$ , nor is 0 assumed to be an annihilator of  $\otimes$ .

**Definition I.1** (Associative Array). An *associative array* is a map  $\mathbf{A} : K_1 \times K_2 \rightarrow V$ , where  $K_1$  and  $K_2$  are finite totally-ordered sets, referred to as *key sets* and whose elements are called *keys*, and  $V$  is the value set.

**Definition I.2** (Transpose). If  $\mathbf{A} : K_1 \times K_2 \rightarrow V$  is an associative array, then  $\mathbf{A}^\top : K_2 \times K_1 \rightarrow V$  is the associative array defined as

$$\mathbf{A}^\top(k_2, k_1) = \mathbf{A}(k_1, k_2)$$

where  $k_1 \in K_1$  and  $k_2 \in K_2$ .

**Definition I.3** (Array Multiplication). Multiplication of associative arrays is defined as

$$\mathbf{C} = \mathbf{A} \oplus \otimes \mathbf{B} = \mathbf{A} \mathbf{B}$$

or more specifically

$$\mathbf{C}(k_1, k_2) = \bigoplus_{k_3} \mathbf{A}(k_1, k_3) \otimes \mathbf{B}(k_3, k_2)$$

where  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$  are associative arrays

$$\begin{aligned} \mathbf{A} &: K_1 \times K_3 \rightarrow V \\ \mathbf{B} &: K_3 \times K_2 \rightarrow V \\ \mathbf{C} &: K_1 \times K_2 \rightarrow V \end{aligned}$$

and  $k_1 \in K_1$ ,  $k_2 \in K_2$ ,  $k_3 \in K_3$ .

**Definition I.4** (Incidence Arrays). If  $G$  is a graph with vertex set  $K_{\text{out}} \cup K_{\text{in}}$  and edge set  $K$ , then

$\mathbf{E}_{\text{out}} : K \times K_{\text{out}} \rightarrow V$  is a *source incidence array* if  $\mathbf{E}_{\text{out}}(k, a) \neq 0$  if and only if the edge  $k \in K$  is directed outward from the vertex  $a \in K_{\text{out}}$   
 $\mathbf{E}_{\text{in}} : K \times K_{\text{in}} \rightarrow V$  is a *target incidence array* if  $\mathbf{E}_{\text{in}}(k, a) \neq 0$  if and only if the edge  $k \in K$  is directed into the vertex  $a \in K_{\text{in}}$ .

**Definition I.5** (Adjacency Array). If  $G$  is a graph with vertex set  $K_{\text{out}} \cup K_{\text{in}}$  and edge set  $K$ , then  $\mathbf{A} : K_{\text{out}} \times K_{\text{in}} \rightarrow V$  is a *adjacency array* if  $\mathbf{A}(a, b) \neq 0$  if and only if there is an edge with source  $a$  and target  $b$ .

## II. ADJACENCY ARRAY CONSTRUCTION

If  $\mathbf{A}$  is an adjacency array for a graph  $G = (K_{\text{out}} \cup K_{\text{in}}, K)$ , then  $\mathbf{A}(a, b) \neq 0$  if and only if there is an edge  $k$  with source  $a$  and target  $b$ , i.e. so that  $\mathbf{E}_{\text{out}}(k, a) \neq 0$  and  $\mathbf{E}_{\text{in}}(k, b) \neq 0$ . In the case where the product of two non-zero values is non-zero, this can be subsumed to say that  $\mathbf{A}(a, b) \neq 0$  if and only if  $\mathbf{E}_{\text{out}}(k, a) \mathbf{E}_{\text{in}}(k, b)$ . Writing this as

$$\mathbf{E}_{\text{out}}(k, a) \mathbf{E}_{\text{in}}(k, b) = \mathbf{E}_{\text{out}}^\top(a, k) \mathbf{E}_{\text{in}}(k, b)$$

This latter expression looks like a term in the evaluation

$$(\mathbf{E}_{\text{out}}^\top \mathbf{E}_{\text{in}})(a, b) = \bigoplus_{k \in K} \mathbf{E}_{\text{out}}^\top(a, k) \mathbf{E}_{\text{in}}(k, b)$$

but the introduction of more terms means that more assumptions need to be made about the relationships between  $\oplus$ ,  $\otimes$ , and 0.

**Theorem II.1.** Let  $V$  be a set with closed binary operations  $\oplus, \otimes$  with identities  $0, 1 \in V$ . Then the following are equivalent:

- 1)  $\oplus$  and  $\otimes$  satisfy the properties (a)
  - a) *Zero-Sum-Free*:  $a \oplus b = 0$  if and only if  $a = b = 0$ ,
  - b) *No Zero Divisors*:  $a \otimes b = 0$  if and only if  $a = 0$  or  $b = 0$ , and
  - c) *0 is Annihilator for  $\otimes$* :  $a \otimes 0 = 0 \otimes a = 0$ .
- 2) If  $G$  is a graph with out-vertex and in-vertex incidence arrays  $\mathbf{E}_{\text{out}} : K \times K_{\text{out}} \rightarrow V$  and  $\mathbf{E}_{\text{in}} : K \times K_{\text{in}} \rightarrow V$ , then  $\mathbf{E}_{\text{out}}^\top \mathbf{E}_{\text{in}}$  is an adjacency array for  $G$ .

*Proof:* Let  $\mathbf{A} = \mathbf{E}_{\text{out}}^\top \mathbf{E}_{\text{in}}$ .

As above, for  $\mathbf{A}$  to be the adjacency array of  $G$ , the entry  $\mathbf{A}(k_{\text{out}}, k_{\text{in}})$  must be nonzero if and only if there is an edge from  $k_{\text{out}}$  to  $k_{\text{in}}$ , which is equivalent to saying that the entry must be nonzero if and only if there is a  $k \in K$  such that

$$\begin{aligned} \mathbf{E}_{\text{out}}^\top(k_{\text{out}}, k) &\neq 0 \\ \mathbf{E}_{\text{in}}(k, k_{\text{in}}) &\neq 0 \end{aligned}$$

Taken altogether, the above pair of equations imply

$$\bigoplus_{k \in K} \mathbf{E}_{\text{out}}^\top(k_{\text{out}}, k) \otimes \mathbf{E}_{\text{in}}(k, k_{\text{in}}) \neq 0$$

$$\iff \exists k \in K \text{ so that } \mathbf{E}_{\text{out}}^\top(k_{\text{out}}, k) \neq 0 \text{ and } \mathbf{E}_{\text{in}}(k, k_{\text{in}}) \neq 0$$

First, the above condition can be restated in a form that more easily provides the zero-sum-freeness of  $\oplus$ , lack of zero-divisors for  $\otimes$ , and the fact that 0 annihilates under  $\otimes$ . Equation 1 is equivalent to

$$\begin{aligned} \bigoplus_{k \in K} \mathbf{E}_{\text{out}}(k, x) \otimes \mathbf{E}_{\text{in}}(k, y) &= 0 \iff \\ \nexists k \in K \text{ so that } \mathbf{E}_{\text{out}}(k, x) &\neq 0 \text{ and } \mathbf{E}_{\text{in}}(k, y) \neq 0 \quad (1) \end{aligned}$$

which in turn is equivalent to

$$\bigoplus_{k \in K} \mathbf{E}_{\text{out}}(k, x) \otimes \mathbf{E}_{\text{in}}(k, y) = 0 \iff \forall k \in K, \mathbf{E}_{\text{out}}(k, x) = 0 \text{ or } \mathbf{E}_{\text{in}}(k, y) = 0 \quad (2)$$

This expression may be split up into two conditional statements

$$\bigoplus_{k \in K} \mathbf{E}_{\text{out}}(k, x) \otimes \mathbf{E}_{\text{in}}(k, y) = 0 \Rightarrow \forall k \in K, \mathbf{E}_{\text{out}}(k, x) = 0 \text{ or } \mathbf{E}_{\text{in}}(k, y) = 0 \quad (3)$$

and

$$\forall k \in K, \mathbf{E}_{\text{out}}(k, x) = 0 \text{ or } \mathbf{E}_{\text{in}}(k, y) = 0 \Rightarrow \bigoplus_{k \in K} \mathbf{E}_{\text{out}}(k, x) \otimes \mathbf{E}_{\text{in}}(k, y) = 0 \quad (4)$$

**Lemma II.2.** Equation 3 implies that  $V$  is zero-sum-free.

*Proof:* Suppose there exist nonzero  $v, w \in V$  such that  $v \oplus w = 0$ , or that nontrivial additive inverses exist. Then it is possible to choose a graph  $G$  to have edge set  $\{k_1, k_2\}$  and vertex set  $\{a, b\}$ , where both  $k_1, k_2$  start from  $a$  and end at  $b$ . Then defining

$$\begin{aligned} \mathbf{E}_{\text{out}}(k_1, a) &= v \\ \mathbf{E}_{\text{out}}(k_2, a) &= w \\ \mathbf{E}_{\text{in}}(k_i, b) &= 1 \end{aligned}$$

provides proper out-vertex and in-vertex incidence arrays for  $G$ . Moreover, it is the case that

$$\mathbf{E}_{\text{out}}^T \mathbf{E}_{\text{in}}(b, a) = (v \otimes 1) \oplus (w \otimes 1) = v \oplus w = 0$$

which contradicts Equation 3. Therefore, no such nonzero  $v$  and  $w$  may be present in  $V$ , meaning it is necessary that  $V$  be zero-sum-free. ■

**Lemma II.3.** Equation 3 implies that  $V$  has no zero-divisors.

*Proof:* Suppose  $v \otimes w = 0$ . Define the graph  $G$  to have edge set  $\{k\}$  and vertex set  $\{a\}$  with a single self-loop given by  $k$ . Then define

$$\begin{aligned} \mathbf{E}_{\text{out}}(k, a) &= v \\ \mathbf{E}_{\text{in}}(k, a) &= w \end{aligned}$$

to obtain out-vertex and in-vertex incidence arrays for  $G$ . Then

$$\mathbf{E}_{\text{out}}^T \mathbf{E}_{\text{in}}(a, a) = \mathbf{E}_{\text{out}}(k, a) \otimes \mathbf{E}_{\text{in}}(k, a) = v \otimes w = 0$$

Thus, Equation 3 implies that  $v = w = 0$ , and hence  $V$  has no zero-divisors. ■

**Lemma II.4.** Equation 3 implies that 0 annihilates  $V$  under  $\otimes$ .

*Proof:* Suppose  $v \in V$ . Define the graph  $G$  to have edge set  $\{k_1, k_2\}$  and vertex set  $\{a, b\}$ , with self-loops at  $a$  and  $b$  given by  $k_1$  and  $k_2$ , respectively. Define

$$\mathbf{E}_{\text{out}}(k_1, a) = v = \mathbf{E}_{\text{in}}(k_1, a)$$

and

$$\mathbf{E}_{\text{out}}(k_2, b) = v = \mathbf{E}_{\text{in}}(k_2, b)$$

(and all other entries in  $\mathbf{E}_{\text{out}}$  and  $\mathbf{E}_{\text{in}}$  equal to 0) results in out-vertex and in-vertex incidence arrays of  $G$ . Moreover, it is true that

$$\begin{aligned} 0 &= \mathbf{E}_{\text{out}}^T \mathbf{E}_{\text{in}}(a, b) \\ &= \mathbf{E}_{\text{out}}(k_1, a) \otimes \mathbf{E}_{\text{in}}(k_1, b) \oplus \mathbf{E}_{\text{out}}(k_2, a) \otimes \mathbf{E}_{\text{in}}(k_2, b) \\ &= (v \otimes 0) \oplus (0 \otimes v) \end{aligned}$$

By Lemma II.2,  $V$  is zero-sum-free so it follows that  $v \otimes 0 = 0 \otimes v = 0$ . Thus, 0 is an annihilator for  $\otimes$ . ■

Now Theorem II.1(i) is shown to be sufficient for Theorem II.1(ii) to hold. Assume that zero is an annihilator,  $V$  is zero-sum-free, and  $V$  has no zero-divisors. Zero-sum-freeness and the nonexistence of zero divisors gives

$$\begin{aligned} \exists k \in K \text{ so that } \mathbf{E}_{\text{out}}(k, x) \neq 0 \text{ and } \mathbf{E}_{\text{in}}(k, y) \neq 0 \Rightarrow \\ \bigoplus_{k \in K} \mathbf{E}_{\text{out}}(k, x) \otimes \mathbf{E}_{\text{in}}(k, y) \neq 0 \quad (5) \end{aligned}$$

which is the contrapositive of Equation 3. And, that zero is an annihilator gives

$$\begin{aligned} \forall k \in K, \mathbf{E}_{\text{out}}(e, x) = 0 \text{ or } \mathbf{E}_{\text{in}}(e, y) = 0 \Rightarrow \\ \bigoplus_{k \in K} \mathbf{E}_{\text{out}}(k, x) \otimes \mathbf{E}_{\text{in}}(k, y) = 0 \quad (6) \end{aligned}$$

which is (4). As Equation 3 and Equation 4 combine to form Equation 1, it is established that the conditions are sufficient for Equation 1. ■

### III. ADJACENCY ARRAY OF REVERSE GRAPH

The remaining product of the incidence arrays that is defined is  $\mathbf{E}_{\text{in}}^T \mathbf{E}_{\text{out}}$ . The above requirements will now be shown to be necessary and sufficient for the remaining product to be the adjacency array of the reverse of the graph. Recall that the reverse of  $G$  is the graph  $\bar{G}$  in which all the arrows in  $G$  have been reversed. Let  $G$  be a graph with incidence matrices  $\mathbf{E}_{\text{out}}$  and  $\mathbf{E}_{\text{in}}$ .

**Corollary III.1.** Condition (i) in Theorem II.1 are necessary and sufficient so that  $\mathbf{E}_{\text{in}}^T \mathbf{E}_{\text{out}}$  is an adjacency matrix of the reverse of  $G$ .

*Proof:* Let  $\bar{G}$  denote the reverse of  $G$ , and let  $\bar{\mathbf{E}}_{\text{out}}$  and  $\bar{\mathbf{E}}_{\text{in}}$  be out-vertex and in-vertex incidence arrays for  $\bar{G}$ , respectively. Recall that  $\bar{G}$  is defined to have the same edge and vertex sets as  $G$  but changes the directions of the edges, in other words, if an edge  $k$  leaves a vertex  $a$  in  $G$ , then it enters  $a$  in  $\bar{G}$ , and vice versa. As such,  $\mathbf{E}_{\text{out}}(k, a) \neq 0$  if

and only if  $\bar{\mathbf{E}}_{\text{in}}(k, a) \neq 0$ , and likewise  $\mathbf{E}_{\text{in}}(k, a) \neq 0$  if and only if  $\bar{\mathbf{E}}_{\text{out}}(k, a) \neq 0$ . As such, choosing  $\mathbf{E}_{\text{out}} = \bar{\mathbf{E}}_{\text{in}}$  and  $\mathbf{E}_{\text{in}} = \bar{\mathbf{E}}_{\text{out}}$  gives valid in-vertex and out-vertex incidence matrices for  $\bar{G}$ , respectively. Then by Theorem II.1 it can be shown that

$$\bar{\mathbf{E}}_{\text{out}}^T \bar{\mathbf{E}}_{\text{in}} = \mathbf{E}_{\text{in}}^T \mathbf{E}_{\text{out}}$$

■

It is now straightforward to identify algebraic structures that comply with the established criteria. Notably, all zero-sum-free semirings with no zero-divisors comply, such as  $\mathbb{N}$  or  $\mathbb{R}_{\geq 0}$  with the standard addition and multiplication. In addition, any linearly ordered set with  $\oplus$  and  $\otimes$  given by max and min, respectively. Some non-examples, however, include the max-plus algebra or non-trivial Boolean algebras, which do not satisfy the zero-product property, or rings, which except for the zero ring are not zero-sum-free. Furthermore, the value sets of associative arrays need not be defined exclusively as semirings, as several semiring-like structures satisfy the criteria. These structures may lack the properties of additive or multiplicative commutativity, additive or multiplicative associativity, or distributivity of multiplication over addition, which are not necessary to ensure that the product of incidence arrays yields an adjacency array.

The criteria guarantee an accurate adjacency array for any dataset that satisfies them, regardless of value distribution in the incidence arrays. However, if the incidence arrays are known to possess a certain structure, it is possible to circumvent some of the conditions and still always produce adjacency arrays. For example, if each key set of an undirected incidence array  $\mathbf{E}$  is a list of documents and the array entries are sets of words shared by documents, then it is necessary that a word in  $\mathbf{E}(i, j)$  and  $\mathbf{E}(m, n)$  has to be in  $\mathbf{E}(i, n)$  and  $\mathbf{E}(m, j)$ . This structure means that when multiplying  $\mathbf{E}^T \mathbf{E}$  using  $\oplus = \cup$  and  $\otimes = \cap$ , a nonempty set will never be “multiplied” by (intersected with) a disjoint nonempty set. This eliminates the need for the zero-product property to be satisfied, as every multiplication of nonempty sets is already guaranteed to produce a nonempty set. The array produced will contain as entries a list of words shared by those two documents.

Though the criteria ensure that the product of incidence arrays will be an adjacency array, they do not ensure that certain matrix properties hold. For example, the property  $(\mathbf{A}\mathbf{B})^T = \mathbf{B}^T \mathbf{A}^T$  may be violated under these criteria, as  $(\mathbf{E}_{\text{out}}^T \mathbf{E}_{\text{in}}^T)$  is not necessarily equal to  $\mathbf{E}_{\text{in}}^T \mathbf{E}_{\text{out}}$ . (For this matrix transpose property to always hold, the operation  $\otimes$  would have to be commutative.)

#### IV. GRAPH CONSTRUCTION WITH DIFFERENT SEMIRINGS

The ability to change  $\oplus$  and  $\otimes$  operations allows different graph adjacency arrays to be constructed using the same

element-wise addition, element-wise multiplication, and array multiplication syntax. Specific pairs of operations are best suited for constructing certain types of adjacency arrays. The pattern of edges resulting from array multiplication of incidence arrays is generally preserved for various semirings. However, the non-zero values assigned to the edges can be very different and enable the construction different graphs.

For example, constructing an adjacency array of the graph of music writers connected to music genres from Figure 1 begins with selecting the incidence sub-arrays  $\mathbf{E}_1$  and  $\mathbf{E}_2$  as shown in Figure 2. Array multiplication of  $\mathbf{E}_1^T$  with  $\mathbf{E}_2$  produces the desired adjacency array of the graph. Figure 3 illustrates this array multiplication for different operator pairs  $\oplus$  and  $\otimes$ .

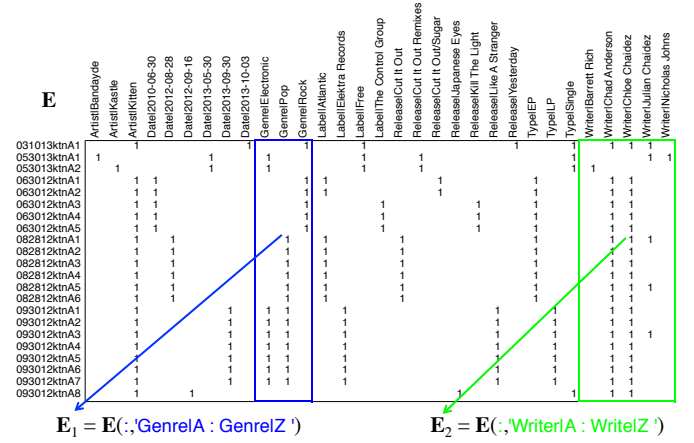


Figure 1. D4M sparse associative array  $\mathbf{E}$  representation of a table of data from a music database. The column key and the value are concatenated with a separator symbol (in this case |) resulting in every unique pair of column and value having its own column in the sparse view. The new value is usually 1 to denote the existence of an entry. Column keys are an ordered set of database fields. Sub-arrays  $\mathbf{E}_1$  and  $\mathbf{E}_2$  are selected with Matlab-style notation to denote all of the row keys and ranges of column keys.

The pattern of edges among vertices in the adjacency arrays shown Figure 3 are the same for the different operator pairs, but the edge weights differ. All the non-zero values in  $\mathbf{E}_1$  and  $\mathbf{E}_2$  are 1. All the  $\otimes$  operators in Figure 3 have the property

$$0 \otimes 1 = 1 \otimes 0 = 0$$

for their respective values of zero be it 0,  $-\infty$ , or  $\infty$ . Likewise, all the  $\otimes$  operators in Figure 3 also have the property

$$1 \otimes 1 = 1$$

except where  $\otimes = +$ , in which case

$$1 \otimes 1 = 2$$

The differences in the adjacency array weights are less pronounced then if the values of  $\mathbf{E}_1$  and  $\mathbf{E}_2$  were more diverse. The most apparent difference is between the  $+. \times$

$E_1$	Genre Electronic	Genre Pop	Genre Rock
031013ktnA1			1
053013ktnA1	1		
053013ktnA2	1		
063012ktnA1			1
063012ktnA2			1
063012ktnA3			1
063012ktnA4			1
063012ktnA5			1
082812ktnA1		1	
082812ktnA2		1	
082812ktnA3		1	
082812ktnA4		1	
082812ktnA5		1	
082812ktnA6		1	
093012ktnA1	1	1	
093012ktnA2	1	1	
093012ktnA3	1	1	
093012ktnA4	1	1	
093012ktnA5	1	1	
093012ktnA6	1	1	
093012ktnA7	1	1	

$E_2$	Writer Barrett Rich	Writer Chad Anderson	Writer Chloe Chadez	Writer Julian Chadez	Writer Nicholas Johns
031013ktnA1		1	1	1	
053013ktnA1				1	1
053013ktnA2	1				
063012ktnA1		1	1		
063012ktnA2		1	1		
063012ktnA3		1	1		
063012ktnA4		1	1		
063012ktnA5		1	1		
082812ktnA1		1	1	1	
082812ktnA2		1	1		
082812ktnA3		1	1		
082812ktnA4		1	1		
082812ktnA5		1	1	1	
082812ktnA6		1	1		
093012ktnA1	1	1			
093012ktnA2	1	1			
093012ktnA3	1	1	1		
093012ktnA4	1	1			
093012ktnA5	1	1			
093012ktnA6	1	1			
093012ktnA7	1	1			
093012ktnA8	1	1			

Figure 2. Incidence arrays of music writers and music genres  $E_1$  and  $E_2$  as defined in Figure 1 for different tracks of music.

semiring and the other semirings in Figure 3. In the case of  $+. \times$  semiring, the  $\oplus$  operation  $+$  aggregates values from all the edges between two vertices. Additional positive edges will increase the overall weight in the adjacency array. In the other pairs of operations, the  $\oplus$  operator is either max or min, which effectively selects only one edge weight to use for assigning the overall weight. Additional edges will only impact the edge weight in the adjacency array if the new edge is an appropriate maximum or minimum value. Thus,  $+. \times$  constructs adjacency arrays that aggregate all the edges. The other semirings construct adjacency arrays that select extremal edges. Each can be useful for construction graph adjacency arrays in appropriate context.

The impact of different semirings on the graph adjacency array weights are more pronounced if the values of  $E_1$  and  $E_2$  are more diverse. Figure 4 modifies  $E_1$  so that a value of 2 is given to the non-zero values in the column Genre|Pop and a value of 3 is given to the non-zero values in the column Genre|Rock.

Figure 5 shows the results of constructing adjacency arrays with  $E_1$  and  $E_2$  using different semirings. The impact of changing the values in  $E_1$  can be seen by comparing Figure 3 with Figure 5. For the  $+. \times$  semiring, the values in the adjacency array rows Genre|Pop and Genre|Rock are multiplied by 2 and 3. The increased adjacency array values for these rows are a result of the  $\otimes$  operator being arithmetic multiplication  $\times$  so that

$$\begin{aligned} 2 \otimes 1 &= 2 \times 1 = 2 \\ 3 \otimes 1 &= 3 \times 1 = 3 \end{aligned}$$

For the max.+ and min.+ semirings, the values in the

$E_1^T$	$+. \times$	$E_2$	=	Genre Electronic	1	7	7	2	1
				Genre Pop	13	13	3		
				Genre Rock	6	6	1		

$E_1^T$	max.+	$E_2$	=	Genre Electronic	2	2	2	2	2
	min.+			Genre Pop	2	2	2		
				Genre Rock	2	2	2		

$E_1^T$	max.min	$E_2$	=	Genre Electronic	1	1	1	1	1
				Genre Pop	1	1	1		
				Genre Rock	1	1	1		

$E_1^T$	min.max	$E_2$	=	Genre Electronic	1	1	1	1	1
				Genre Pop	1	1	1		
				Genre Rock	1	1	1		

$E_1^T$	max. $\times$	$E_2$	=	Genre Electronic	1	1	1	1	1
	min. $\times$			Genre Pop	1	1	1		
				Genre Rock	1	1	1		

Figure 3. Creating a graph of music writers related to music genres can be computed by multiplying  $E_1$  and  $E_2$  as defined in Figure 2. This correlation is performed using the transpose operation  $^T$  and the array multiplication operation  $\otimes$ . The resulting associative array has row keys taken from the column keys of  $E_1$  and column keys taken from the column keys of  $E_2$ . The values represent the weights on the edges between the vertices of the graph. Different pairs of operations  $\oplus$  and  $\otimes$  produce different results. For display convenience, operator pairs that produce the same values in this specific example are stacked.

adjacency array rows Genre|Pop and Genre|Rock are larger by and 1 and 2. The larger values in the adjacency array of these rows is due to the  $\otimes$  operator being arithmetic addition  $+$  resulting in

$$\begin{aligned} 2 \otimes 1 &= 2 + 1 = 3 \\ 3 \otimes 1 &= 3 + 1 = 4 \end{aligned}$$

For the max.min semiring, Figure 3 and Figure 5 have the same adjacency array because  $E_2$  is unchanged. The  $\otimes$  operator corresponding to the minimum value function continues to select the smaller non-zero values from  $E_2$

$$\begin{aligned} 2 \otimes 1 &= \min(2, 1) = 1 \\ 3 \otimes 1 &= \min(3, 1) = 1 \end{aligned}$$

In contrast, for the min.max semiring, the values in the adjacency array rows Genre|Pop and Genre|Rock are larger by and 1 and 2. The increase in adjacency array values for these rows are a result of the  $\otimes$  operator selecting the larger non-zero values from  $E_1$

$$\begin{aligned} 2 \otimes 1 &= \max(2, 1) = 2 \\ 3 \otimes 1 &= \max(3, 1) = 3 \end{aligned}$$

Finally, for the max. $\times$  and min. $\times$  semirings, the values in the adjacency array rows Genre|Pop and Genre|Rock are increased by and 1 and 2. Similar to the  $+. \times$  semiring, the

$E_1$	GenreElectronic	GenrePop	GenreRock
031013ktnA1			3
053013ktnA1	1		
053013ktnA2	1		
063012ktnA1			3
063012ktnA2			3
063012ktnA3			3
063012ktnA4			3
063012ktnA5			3
082812ktnA1		2	
082812ktnA2		2	
082812ktnA3		2	
082812ktnA4		2	
082812ktnA5		2	
082812ktnA6		2	
093012ktnA1	1	2	
093012ktnA2	1	2	
093012ktnA3	1	2	
093012ktnA4	1	2	
093012ktnA5	1	2	
093012ktnA6	1	2	
093012ktnA7	1	2	

$E_2$	WriterBarrett Rich	WriterChad Anderson	WriterChloe Chadez	WriterJulian Chadez	WriterNicholas Johns
031013ktnA1	1	1	1		
053013ktnA1			1	1	
053013ktnA2	1				
063012ktnA1		1	1		
063012ktnA2		1	1		
063012ktnA3		1	1		
063012ktnA4		1	1		
063012ktnA5		1	1		
082812ktnA1		1	1	1	
082812ktnA2		1	1		
082812ktnA3		1	1		
082812ktnA4		1	1		
082812ktnA5		1	1	1	
082812ktnA6		1	1		
093012ktnA1	1	1			
093012ktnA2	1	1			
093012ktnA3	1	1	1		
093012ktnA4	1	1			
093012ktnA5	1	1			
093012ktnA6	1	1			
093012ktnA7	1	1			
093012ktnA8	1	1			

Figure 4. Incidence arrays from Figure 2 modified so that the non-zero values of  $E_1$  take on the values 1, 2, and 3.

larger adjacency array values for these rows are a result of the  $\otimes$  operator being arithmetic multiplication  $\times$  resulting in

$$2 \otimes 1 = 2 \times 1 = 2$$

$$3 \otimes 1 = 3 \times 1 = 3$$

Figures 3 and 5 show that a wide range of graph adjacency arrays can be constructed via array multiplication of incidence arrays over different semirings. A synopsis of the graph constructions illustrated in Figures 3 and 5 is as follows

- $+. \times$  sum of products of edge weights connecting two vertices; computes the strength of all connections between two connected vertices.
- $\max. \times$  maximum of products edge weights connecting two vertices; selects the edge with largest weighted product of all the edges connecting two vertices.
- $\min. \times$  minimum of products edge weights connecting two vertices; selects the edge with smallest weighted product of all the edges connecting two vertices.
- $\max. +$  maximum of sum of edge weights connecting two vertices; selects the edge with largest weighted sum of all the edges connecting two vertices.
- $\min. +$  minimum of sum of edge weights connecting two vertices; selects the edge with smallest weighted sum of all the edges connecting two vertices.
- $\max. \min$  maximum of the minimum of weights connecting two vertices; selects the largest of all the shortest connections between two vertices.

$E_1^T$	$+. \times$	$E_2$	=	GenreElectronic	1	7	7	2	1
				GenrePop	26	26	6		
				GenreRock	18	18	3		

$E_1^T$	$\max. +$	$E_2$	=	GenreElectronic	2	2	2	2	2
	$\min. +$			GenrePop	3	3	3		
				GenreRock	4	4	4		

$E_1^T$	$\max. \min$	$E_2$	=	GenreElectronic	1	1	1	1	1
				GenrePop	1	1	1		
				GenreRock	1	1	1		

$E_1^T$	$\min. \max$	$E_2$	=	GenreElectronic	1	1	1	1	1
				GenrePop	2	2	2		
				GenreRock	3	3	3		

$E_1^T$	$\max. \times$	$E_2$	=	GenreElectronic	1	1	1	1	1
	$\min. \times$			GenrePop	2	2	2		
				GenreRock	3	3	3		

Figure 5. Building a graph of music writers connected with the music genres can be accomplished by multiplying  $E_1$  and  $E_2$  as defined in Figure 4. The correlation is computed with the transpose operation  $^T$  and array multiplication  $\otimes$ . The resulting associative array has row keys taken from the column keys of  $E_1$  and column keys taken from the column keys of  $E_2$ . The values represent the weights on the edges between the vertices of the graph. Different pairs of operations  $\oplus$  and  $\otimes$  produce different results. For display convenience, operator pairs that produce the same values in this specific example are stacked.

$\min. \max$  minimum of the maximum of weights connecting two vertices; selects the smallest of all the largest connections between two vertices.

## V. CONCLUSION

Graph construction, a fundamental operation in a data processing pipeline, is typically done by multiplying the incidence array representations of a graph,  $E_{in}$  and  $E_{out}$ , to produce an adjacency array of the graph,  $A$ . The mathematical criteria to determine if  $A$  will have the required structure of the adjacency array of the graph over are as follows. Let  $V$  be a set with closed binary operations  $\oplus, \otimes$  with identities  $0, 1 \in V$ . Then the following are equivalent:

- 1)  $\oplus$  and  $\otimes$  satisfy the properties (a)
  - a) Zero-Sum-Free:  $a \oplus b = 0$  if and only if  $a = b = 0$ ,
  - b) No Zero Divisors:  $a \otimes b = 0$  if and only if  $a = 0$  or  $b = 0$ , and
  - c) 0 is Annihilator for  $\otimes$ :  $a \otimes 0 = 0 \otimes a = 0$ .
- 2) If  $G$  is a graph with out-vertex and in-vertex incidence arrays  $E_{out} : K \times K_{out} \rightarrow V$  and  $E_{in} : K \times K_{out} \rightarrow V$ , then  $E_{out}^T E_{in}$  is an adjacency array for  $G$ .

The values in the resulting adjacency array are determined by the corresponding addition  $\oplus$  and multiplication  $\otimes$  operations used to perform the array multiplication.

# ACKNOWLEDGMENT

The authors would like to thank Paul Burkhardt, Alan Edelman, Sterling Foster, Vijay Gadepally, Sam Madden, Dave Martinez, Tom Mattson, Albert Reuther, Victor Roytburd, and Michael Stonebraker.

# REFERENCES

- [Anderson et al 2016] M. Anderson, N. Sundaram, N. Satish, M. Patwary, T. L. Willke, & P. Dubey, GraphPad: Optimized Graph Primitives for Parallel and Distributed Platforms, submitted
- [Bodin & Kursh 1979] L. Bodin & S. Kursh, *A detailed description of a computer system for the routing and scheduling of street sweepers*, Computers & Operations Research, 6(4), 181-198, 1979
- [Brualdi 1967] R.A. Brualdi, *Kronecker products of fully indecomposable matrices and of ultrastrong digraphs*, Journal of Combinatorial Theory, 2:135-139, 1967
- [Bruck & Ryser 1949] R. Bruck & H. Ryser, *The nonexistence of certain finite projective planes*, Canadian Journal of Mathematics, 1, 88-93, 1949
- [Buluç & Gilbert 2011] A. Buluç & J. Gilbert, The Combinatorial BLAS: Design, implementation, and applications . International Journal of High Performance Computing Applications (IJHPCA), 2011
- [Buluç 2015] A. Buluç, GraphBLAS Special Session, IEEE HPEC 2015, Waltham, MA
- [Dibert et al 2015] K. Dibert, H. Jansen & J. Kepner, *Algebraic Conditions for Generating Accurate Adjacency Arrays*, IEEE MIT Undergraduate Research Technology Conference, 2015
- [Dobrzanskyj & Freudenstein 1967] L. Dobrzanskyj & F. Freudenstein, *Some applications of graph theory to the structural analysis of mechanisms*, Journal of Engineering for Industry, 89(1), 153-158, 1967
- [Ekanadham et al 2014] K. Ekanadham, B. Horn, J. Jann, M. Kumar, J. Moreira, P. Pattnaik, M. Serrano, G. Tanase, H. Yu, Graph Programming Interface: Rationale and Specification, IBM Research Report, RC25508 (WAT1411-052) November 19, 2014
- [Fisher & Wing 1965] G. Fisher & O. Wing, *Computer recognition and extraction of planar graphs from the incidence matrix*, IEEE Transactions on Circuit Theory, 13(2), 154-163, 1966
- [Ford & Fulkerson 1962] L. Ford & D. Fulkerson, *Flows in networks*, Princeton university press, 1962
- [Fulkerson & Gross 1965] D. Fulkerson & O. Gross, *Incidence matrices and interval graphs*, Pacific journal of mathematics, 15(3), 835-855, 1965
- [Harary & Tauth 1964] F. Harary & C.A. Tauth, *Connectedness of products of two directed graphs*, SIAM Journal on Applied Mathematics, 14:250-254, 1966
- [Harary 1969] F. Harary, *Graph Theory*, Reading:Addison-Wesley, 1969
- [Hutchison et al 2015] D. Hutchison, J. Kepner, V. Gadepally, & A. Fuchs, Graphulo implementation of server-side sparse matrix multiply in the Accumulo database, IEEE High Performance Extreme Computing (HPEC) Conference, Waltham, MA, September 2015.
- [Kepner & Gilbert 2011] J. Kepner & J. Gilbert (editors), *Graph Algorithms in the Language of Linear Algebra*, SIAM Press, Philadelphia, 2011
- [Kepner et al 2012] J. Kepner, W. Arcand, W. Bergeron, N. Bliss, R. Bond, C. Byun, G. Condon, K. Gregson, M. Hubbell, J. Kurz, A. McCabe, P. Michaleas, A. Prout, A. Reuther, A. Rosa & C. Yee, Dynamic Distributed Dimensional Data Model (D4M) Database and Computation System, ICASSP (International Conference on Acoustics, Speech, and Signal Processing), 2012, Kyoto, Japan
- [Konig 1931] D. Konig, *Graphen und Matrizen (Graphs and Matrices)*, Matematikai Lapok, 38:116-119, 1931.
- [Konig 1936] D. Konig, *Theorie der endlichen und unendlichen graphen (Theory of finite and infinite graphs)*, Leipzig:Akademie Verlag M.B.H., 1936; see Richard McCourt (Birkhauser 1990) for an english translation of this classic work
- [Mattson et al 2013] T. Mattson, D. Bader, J. Berry, A. Buluç, J. Dongarra, C. Faloutsos, J. Feo, J. Gilbert, J. Gonzalez, B. Hendrickson, J. Kepner, C. Leiserson, A. Lumsdaine, D. Padua, S. Poole, S. Reinhardt, M. Stonebraker, S. Wallach, & A. Yoo, Standards for Graph Algorithms Primitives, IEEE HPEC 2013, Waltham, MA
- [Mattson 2014a] T. Mattson, Workshop on Graph Algorithms Building Blocks, IPDPS 2014, Pheoniz, AZ
- [Mattson 2014b] T. Mattson, GraphBLAS Special Session, IEEE HPEC 2014, Waltham, MA
- [Mattson 2015] T. Mattson, Workshop on Graph Algorithms Building Blocks, IPDPS 2015, Hyderabad, India
- [Mattson 2016] T. Mattson, Workshop on Graph Algorithms Building Blocks, IPDPS 2016, Chicago, IL
- [McAndrew 1963] M.H. McAndrew, *On the product of directed graphs*, Proceedings of the American Mathematical Society, 14:600-606, 1963
- [McAndrew 1965] M.H. McAndrew, *On the polynomial of a directed graph*, Proceedings of the American Mathematical Society, 16:303-309, 1965
- [Sabadusi 1960] G. Sabadusi, *Graph multiplication*, Mathematische Zeitschrift, 72:446-457, 1960
- [Tarjan 1972] R. Tarjan, *Depth-first search and linear graph algorithms*, SIAM journal on computing, 1(2), 146-160, 1972
- [Teh & Yap 1964] H.H. Teh & H.D. Yap, *Some construction problems of homogeneous graphs*, Bulletin of the Mathematical Society of Nanyang University, 164-196, 1964

[Weischel 1962] P.M. Weischel. The Kronecker product of graphs.  
*Proceedings of the American Mathematical Society*, 13(1):47–  
52, 1962

[Zhang et al 2016] P. Zhang, M. Zalewski, A. Lumsdaine, S.  
Misurda, & S. McMillan, GBTL-CUDA: Graph Algorithms  
and Primitives for GPUs, GABB workshop at IPDPS 2016